

Vocabulary as a Core Public Health Infrastructure Component

Cecil O. Lynch, MD, MS
UC Davis/CADHS
colynch@ucdavis.edu
PHIN 2005

HIE Study Savings

Dr. Middleton's talk

- 337 Billion saved with vocabulary
- 78 Billion saved with messaging but without vocabulary
- 259 Billion – The value of this talk

Overview

- Vocabulary Basics
- CHI vocabularies
- Terminology Role in Infrastructure
- Terminology Development
- Terminology Maintenance
- Terminology Distribution
- Relating vocabulary to the data model

The Basics

- Concepts
- Data elements
- Metadata
- Structural vocabulary
- Descriptive vocabulary
- Data Types
- Value domains
- Value sets
- Vocabulary Information Model

Concepts

- Concept defines a unitary mental representation of a real or abstract thing; an atomic unit of thought
- Should be unique in a given terminology
- May have synonyms
- May be a primitive or compositional term

Data Elements

- A single unit of data
- Corresponds to a field in a data base record
- The real world instantiation of a concept
- A textbox entry on your web form

Metadata

- Data about a datum
- Allows for a full description of a data element such that the data element can be classified and potentially reproduced
- Provides the necessary information to allow vocabulary interoperability
- Several representation schemes, ex. Dublin Core, W3C SKOS (Simple Knowledge Organization System)

UMLS Metadata

Name

Role

Concrete

Documentation

<html>
<head>

</head>
<body>

Constraints

Template Slots

Name	Cardinality	Type	Other Facets
broader	multiple	String	
cui	single	String	
documentation	single	String	
name	single	String	
narrower	multiple	String	
related_slot	multiple	String	
semantic_type	multiple	String	
synonym	multiple	String	
translation	multiple	Instance of Condition_metaClass or KB	

Cui

Broader

Arthropod-borne infectious disease (disorder)
Dengue
Togaviridae Infections

Narrower

Dengue Hemorrhagic Fever
Dengue Shock Syndrome

Related Slot

Arthropod-borne organism (organism)
Communicable Diseases, Emerging
Dengue Virus
Gradual onset (contextual qualifier) (qualifier value)

Semantic Type

Disease or Syndrome

Synonym

Breakbone fever
Dengue (disorder)
Dengue Fever
Dengue fever [classical dengue]
Dengue fever virus infection
DUENGERO
FEVER, BREAK BONE
FEVER, DANDY

Structural Vocabulary

- Vocabulary meant to define the structural classes in a data model from which objects can be created
- Examples; the HL7 Version 3 structural codes that define Class Codes, Mood codes etc. or a high level namespace in an ontology model describing a grouping for more primitive concepts such as “Living Organism” with sub-classes of “Virus, Bacterium, Fungi, Parasite”
- Corresponds to the column name in a database table

XML Structural Vocabulary

```
<xsd:complexType name="CD">
  <xsd:complexContent>
    <xsd:extension base="ANY">
      <xsd:sequence>
        <xsd:element name="originalText" type="ED" ... />
        <xsd:element name="qualifier" type="CR" ... />
        <xsd:element name="translation" type="CD" ... />
      </xsd:sequence>
      <xsd:attribute name="code" type="cs" ... />
      <xsd:attribute name="codeSystem" type="uid" ... />
      <xsd:attribute name="codeSystemName" type="st" ... />
      <xsd:attribute name="codeSystemVersion" type="st" ... />
      <xsd:attribute name="displayName" type="st" ... />
    </xsd:extension>
  </xsd:complexContent>
</xsd:complexType>
```

Descriptive Vocabulary

- Defines the terms and codes for concepts
- The data used to populate a data element
- Provides the core component for system and data interoperability
- Examples: Anthrax; Gram negative rod; centrifugal rash; MRI of brain
- Corresponds to an entry in a database column field

Data Types

- Constrain the values for a data element
- Provide concept level semantics
- Do not have state or commutable properties (are immutable)
- Examples: Integer; Float; Boolean; HL7 CD, CE, CV, CS, CR Data Types

HL7 CD (Concept Descriptor)

Table 11: Property Summary of Concept Descriptor

Name	Type	Description
code	ST	The plain code symbol defined by the code system. For example, "784.0" is the code symbol of the ICD-9 code "784.0" for headache.
codeSystem	UID	Specifies the code system that defines the code.
codeSystemName	ST	The common name of the coding system.
codeSystemVersion	ST	If applicable, a version descriptor defined specifically for the given code system.
displayName	ST	A name or title for the code, under which the sending system shows the code value to its users.
originalText	ED	The text or phrase used as the basis for the coding.
translation	SET<CD>	A set of other concept descriptors that translate this concept descriptor into other code systems.
qualifier	LIST<CR>	Specifies additional codes that increase the specificity of the primary code.

Value Domains

- Provides a high level grouping for all things possible in a given domain
- Example: HL7 EntityClass describes all entities used in HL7 V3 messaging including Person, Microorganism, Manufactured material
- Example: SNOMED CT or LOINC

CHI Vocabularies

Anatomy	SNOMED CT ; NCI Thesaurus
Demographics	HL7 2.4 +
Diagnosis	SNOMED CT
Procedures (non-lab)	SNOMED CT
Procedures (laboratory)	LOINC
Lab Results	SNOMED CT
Immunizations	HL7 2.3.1 + (CVX, MVX)

CHI Vocabularies

Clinical Drugs	RxNorm
Drug Classifications	National Drug File Reference Terminology
Manufactured Dosage Form	FDA/CDER
Drug Product	FDA NDC
Nursing	SNOMED CT
Text Based Reports	HL7 CDA Release 1.0
Multimedia in Reports	None decided yet

Value Sets

- Restricts a Value Domain
- Ideally contains values from a single coding system, but not necessarily
- Example: The SNOMED CT Organisms that correspond to Nationally Reportable Disease infectious agents
- Example: All LOINC Tests for Brucellosis

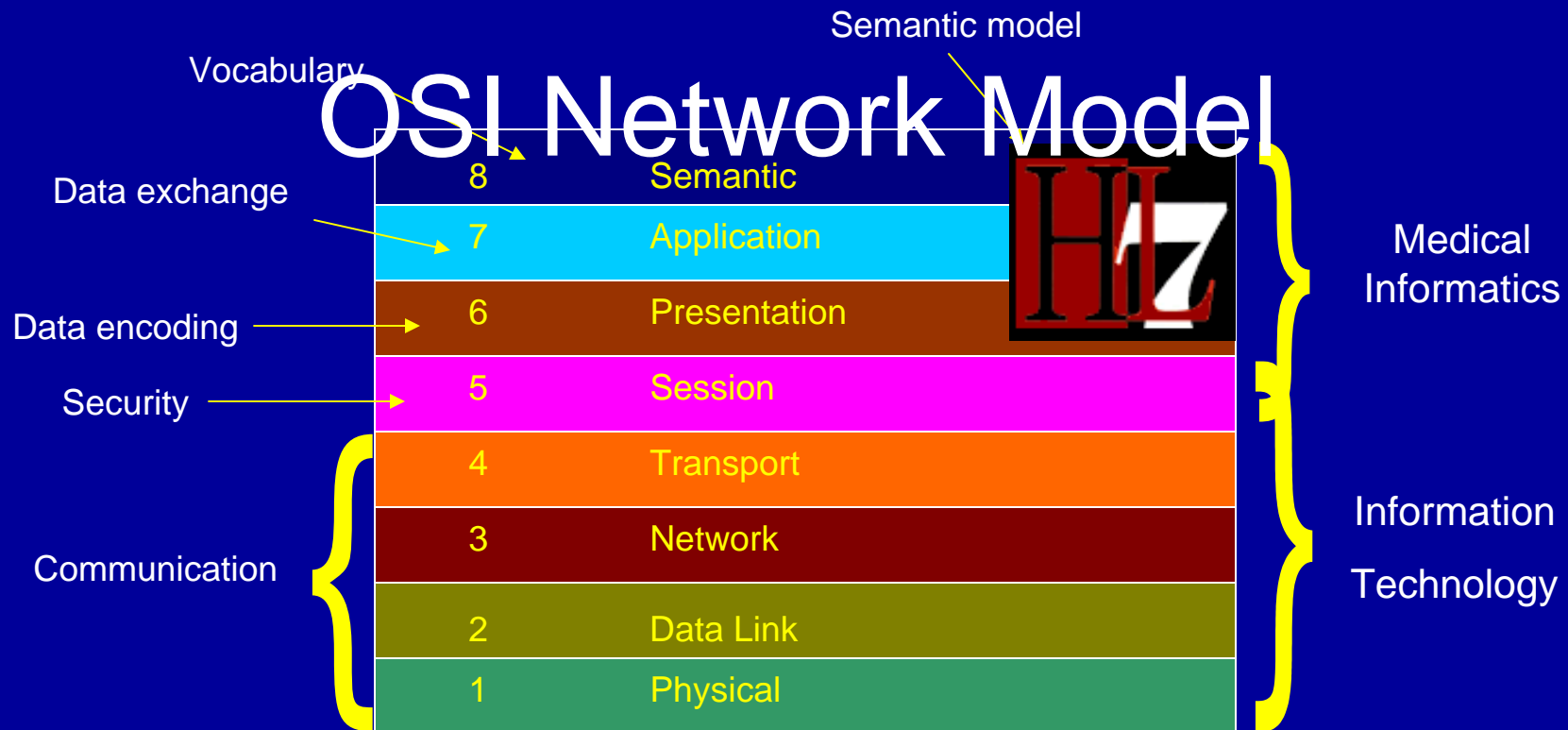
Vocabulary Information Model

- Organize value sets into relational groupings
- Example: All organisms, lab and clinical observations and, procedures related to a particular disease
- Composed of multiple value sets across one or more vocabulary domains

Interoperability Requirements

- Equivalent Data Types
 - Derived from the data model
- Common terminology
 - Unambiguous coding
- Agreed upon format
 - Must be readable from one application to another
- Appropriate connection
 - Message must be routed appropriately across an enterprise

Interoperability Layers



Concept for slide from Stan Huff and Ted Klein

What is the Role of Vocabulary?

- Defines the meaning of data – i.e. changes data to information through instantiation of semantic rules
- Is the Human readable value a user sees
- Allows for intersystem interoperability by disambiguation of the message payload
- Required for data translation
- Required for data aggregation
- Is the single most important component for interoperability

Why Code Data?

- Cold
 - February is a 451893009 month.
 - She meets his gaze with 235346001 stare.
 - Julia is in bed with a 621272006.

Data Aggregation

Term	Description ID	Concept ID
myocardial infarction	37436014	22298006
cardiac infarction	37442013	22298006
heart attack	37443015	22298006
myocardial infarct	1784873012	22298006
MI - Myocardial infarction	1784872019	22298006
infarction of heart	37441018	22298006

Vocabulary Editing and Authoring

- Apelon
 - Commercial
- Health Language
 - Commercial
- LexGrid
 - Open Source
- NCI caDSR
 - Government use

LexGrid Online - Eclipse Platform

File Edit Navigate Search Project Run Window Help

Classic Search

Text of assigned properties Contains of breast

- Other specified disorders of breast: 611.8 (ICD-9/Online)
- Papillary part of fibrous mass of breast: 67774 (FMA-Online/Online)
- Peri-acinar part of fibrous mass proper of breast: 62105 (FMA-Online/Online)
- Periductal lymphatic plexus of breast: 67751 (FMA-Online/Online)
- Periductal part of fibrous mass of breast: 62106 (FMA-Online/Online)
- Perilobular lymphatic plexus of breast: 67752 (FMA-Online/Online)
- Phlebitis and thrombophlebitis of other sites: 451.89 (ICD-9/Online)
- Quadrant of breast: 61373 (FMA-Online/Online)
- Reproductive History: D017584 (MeSH/Online)
- Secondary malignant neoplasm of breast: 198.81 (ICD-9/Online)
- Secondary malignant neoplasm of skin: 198.2 (ICD-9/Online)
- Set of suspensory ligaments of breast: 71433 (FMA-Online/Online)
- Skin of breast: 10464 (FMA-Online/Online)
- Skin of breast proper: 58058 (FMA-Online/Online)
- Skin of lower quadrant of breast: 61425 (FMA-Online/Online)
- Skin of quadrant of breast: 61419 (FMA-Online/Online)
- Skin of subdivision of breast: 61418 (FMA-Online/Online)
- Skin of upper quadrant of breast: 61424 (FMA-Online/Online)
- Solitary cyst of breast: 610.0 (ICD-9/Online)
- Specified congenital anomalies of breast: 757.6 (ICD-9/Online)
- Subareolar part of fibrous mass of breast: 62107 (FMA-Online/Online)
- Subdivision of breast: 62112 (FMA-Online/Online)
- Subdivision of lymphatic plexus of breast: 73224 (FMA-Online/Online)
- Subdivision of superficial fascia of breast: 19892 (FMA-Online/Online)
- Superficial fascia of breast: 19874 (FMA-Online/Online)
- Superficial fascia of breast proper: 73132 (FMA-Online/Online)
- Superficial fascia of lower quadrant of breast: 61630 (FMA-Online/Online)
- Superficial fascia of quadrant of breast: 61628 (FMA-Online/Online)
- Superficial fascia of upper quadrant of breast: 61629 (FMA-Online/Online)
- Surface of lower quadrant of breast: 61456 (FMA-Online/Online)
- Surface of quadrant of breast: 61448 (FMA-Online/Online)
- Surface of skin of breast: 67738 (FMA-Online/Online)
- Surface of upper quadrant of breast: 61449 (FMA-Online/Online)
- Suspensory ligament of breast: 58040 (FMA-Online/Online)
- Tegafur: D005641 (MeSH/Online)
- Upper inner quadrant of breast: 61375 (FMA-Online/Online)
- Upper outer quadrant of breast: 61380 (FMA-Online/Online)
- Weaning: D014886 (MeSH/Online)

Relations Tree

- Upper outer quadrant of breast
 - Detail ...
 - Source for the following relationships ...
 - hasSubtype
 - Target of the following relationships ...
 - Breast > regional-part > Upper outer quadrant of breast
 - fm-live_12276 > related-part > Upper outer quadrant of breast
 - Quadrant of breast > hasSubtype > Upper outer quadrant of breast

Properties

Setting	Value
Coded Entry	
Concept Code	61373
Description	Quadrant of breast
Entity Description	
Other	
Concept Status	
First Version	
Is Active	false
Is Anonymous	false
Last Version	false
Mod Version	

Relations Graph

Rotate

```
graph BT; Breast -- regional-part --> Quadrant_of_breast[Quadrant of breast]; Quadrant_of_breast -- hasSubtype --> Upper_outer[Upper outer quadrant of breast]; Quadrant_of_breast -- hasSubtype --> Upper_inner[Upper inner quadrant of breast]; Quadrant_of_breast -- hasSubtype --> Lower_outer[Lower outer quadrant of breast]; Quadrant_of_breast -- hasSubtype --> Lower_inner[Lower inner quadrant of breast]; Quadrant_of_breast -- hasSubtype --> Subdivision[Subdivision of breast]; style Upper_outer stroke:#0000FF,stroke-width:2px
```

The Entity Description of the Versionable And Describable

Vocabulary Maintenance

Code

4241002

DisplayName

Listeriosis

OriginalText

Listeriosis

CodeSystem

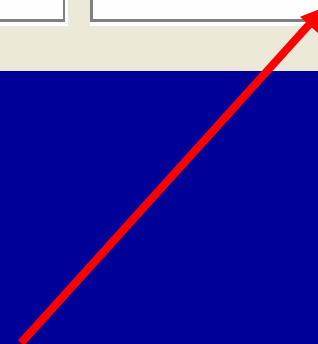
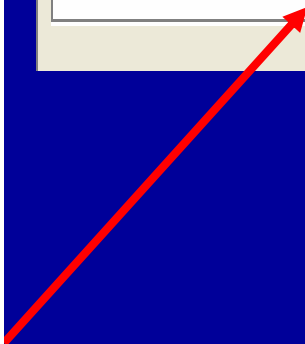
2.16.840.1.113883.6.96

CodeSystemName

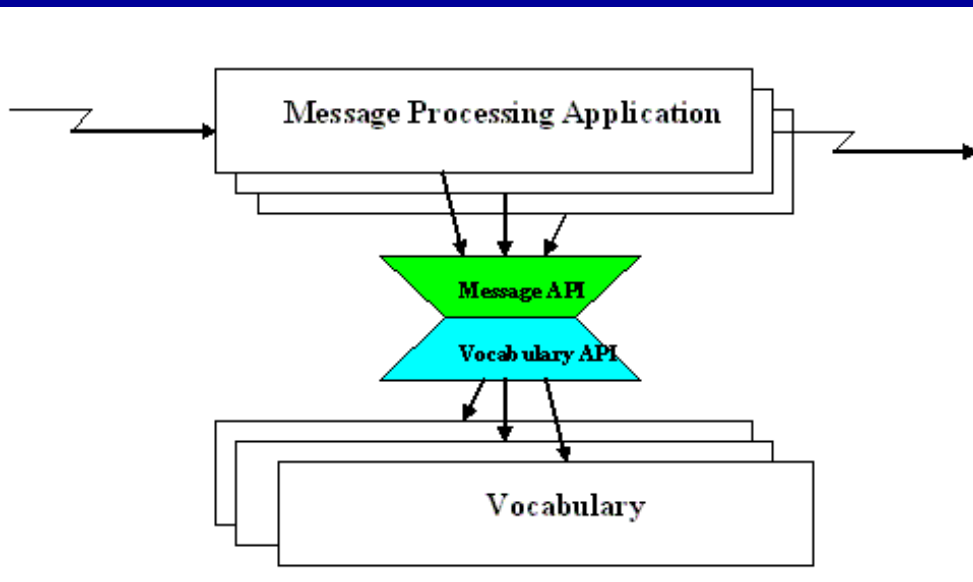
SNOMED CT

CodeSystemVersion

20040731Core



HL7 Common Terminology Service



- Implements a core subset of the OMG TQS specification
- Provides a messaging API and a vocabulary API
- Functions as a runtime message filter
- API's can act as a standalone resource
- Works with LexGrid Editor and binds to Protégé API

PHIN Vocabulary Access and Distribution Services

- PHIN VADS components include an html browser for manual searching, viewing, and download of PHIN approved vocabularies, web services connections for automated functionalities, and a Java application programming interface (API) and data store which can facilitate the development and management of vocabularies within PHIN applications.

PHIN VADS



PHIN Vocabulary Access and Distribution System (VADS)

[VADS Home](#)[PHIN Vocabulary Home](#)[Help / Glossary](#)[CDC Home](#) > [PHIN Home](#) > PHIN VADS

PHIN Vocabulary Access and Distribution System (VADS)

The PHIN Vocabulary Access and Distribution System (PHIN VADS) is a set of tools within the PHIN Vocabulary Services (PHIN VS) that provides a coordinated means for stakeholders to access, distribute, and manage vocabularies. Note: *The PHIN VADS browser has been built to browse, download and search for single code sets and concepts. As such, robust navigation and hierarchical viewing are not available.*

PHIN Recommended

Contains a list of vocabulary concepts commonly used within PHIN (e.g. Race codes, Gender codes).

[PHIN Recommended Vocabularies Home](#)

Application-based

Contains listings of vocabulary concepts associated with PHIN applications (e.g. NEDSS Base System, National Healthcare Safety Network).

Message-based

Contains listings of vocabulary concepts associated with PHIN messages (e.g. NND HL7 V3 Case Notification).

Standard-based Vocabularies

Contains a list of external standardized vocabularies from which PHIN vocabulary listings are drawn (e.g. HL7, LOINC).

Ontological Services

- Can be implemented in RDF or XML
- Tools include Protégé, Clue Browser, Ontoedit
- Queried by a well defined API
- Protégé also can be implemented as a service either through OWL or Java
- Advantage includes the ability to instantiate the Vocabulary Information Model

CLASS BROWSER

For Project: ID_KB_LynchV7YI

Class Hierarchy

Bacterial disease

Aeromonas infection

African tick bite fever

Anthrax

Arizona paracolon gastrointestinal tract infection

Bacterial meningitis

Bartonellosis

Campylobacteriosis

Cat scratch fever

Chancroid

Chlamydial infection

Diphtheria

Escherichia coli, enterohemorrhagic, EHEC

Ehrlichiosis

Enteropathogenic Escherichia coli gastroenteritis

Enteropathogenic Escherichia coli infection

CLASS EDITOR

For Class: Anthrax (instance of PublicHealthCase, internal name is ANTHRAX)

Name

ANTHRAX

Condition

Anthrax

Organism or Toxin

Bacillus anthracis

Diagnostic Lab Tests

BACILLUS ANTHRACIS AB BY AGGL FROM SERUM

BACILLUS ANTHRACIS AB BY CF FROM SERUM

BACILLUS ANTHRACIS AB BY EIA FROM SERUM

BACILLUS ANTHRACIS AB BY HA FROM SERUM

BACILLUS ANTHRACIS AB BY IB FROM SERUM

Transmission Mode

airborne

environmental exposure

fomite

☐ Animal Condition

☒ VPD

ClassCode

CASE

Mood Code

EVN

Incubation Period

Anthrax Incubation

Reportable In these Jurisdictions

California

CDC

Hawaii

PrimarySyndromeGroup

Category A biological agent

SyndromicGroup

Respiratory disease

Skin rash disease

Gastrointestinal disease

BACILLUS ANTHRACIS AB BY EIA FROM SERUM (instance of Procedure_metaclass, internal name is BACILLUS ANTHRACIS AB BY EIA FROM SERUM)

Name

Lynch22863-5LN

Documentation

Constraints

Role

Concrete

Template Slots

Name	Cardinality	Type	Other Facets
code	required ...	String	
codeSystem	single	String	
codeSystemName	single	String	
codeSystemVersion	single	String	
displayName	single	String	

Code

22863-5

DisplayName

B anthracis Ab Ser QI EIA

ClassCode

OBS

CodeSystem

2.16.840.1.113883.6.1

OriginalText

ANTHRACIS AB BY EIA FROM SERUM

MoodCode

EVN

CodeSystemName

LOINC

Status

Translation

Documentation



Class search for “anthracis” in class tree “SNOMED_CT”

```
SlotValueCondition condition =  
    new SlotValueSimpleConditionOnAnySlot(  
        SymbolValueComparator.MATCHES,  
        “anthracis”,  
        false);  
  
Vector linkedSlots = new Vector();  
Slot slot1 = kb.getSlot(":DIRECT-SUBCLASS");  
linkedSlots.add(slot1);  
  
Instance treeRoot = kb.getInstance("SNOMED_CT");  
  
ReferenceSearchIterator search =  
    new ReferenceSearchIterator(  
        treeRoot,  
        linkedSlots,  
        false,  
        condition);  
while (search.hasNext()) {  
    Frame frame = (Frame) search.next();  
    System.out.println(frame.getBrowserText());  
}
```


Data Modeling of Vocabulary

- Cannot divorce vocabulary from the data model
- HL7 V3 data types are recursive
- Flat relational model does not handle the data types
- May generate the data type in the middle ware
- Consumption of the data types requires consideration in the model

HL7 CD Data Type

● Depressed black eschar (instance of Act.Observation, internal name is ID_KB_LynchV7a_Class_6_Post)

Code

87319000

DisplayName

Eschar

OriginalText

Depressed black eschar

CodeSystem

2.16.840.1.113883.6.96

CodeSystemName

SNOMED CT

CodeSystemVersion

20050131 Core

Qualifier

- Black color
- Depressed structure

Translation













ClassCode

● OBS

MoodCode

● EVN

HL7 CE Data Type

Code	DisplayName	ClassCode   
11467-8	B anthracis Ab Ser IB-aCnc	 OBS
CodeSystem	OriginalText	MoodCode   
2.16.840.1.113883.6.1	BACILLUS ANTHRACIS AB BY IB FROM SERUM	 EVN
CodeSystemName	Status	Translation    
LOINC		
CodeSystemVersion	<input type="checkbox"/> Methodless	
2.14		

Summary

- Vocabulary drives an application
- Interoperability depends on vocabulary
- Standards exist and are becoming more refined but there are gaps
- Authoring, Maintenance and Distribution are complex issues
- CDC is tackling the issue with PHIN VADS
- HL7 has defined a standard for terminology services (CTS)